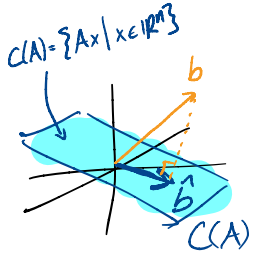
Least Squares Approximation



Let A be an $m \times n$ matrix

Suppose $Ax = b$ has no solution.

This means $b \notin C(A)$

Recall, $\hat{b} = \text{proj}_{C(A)} b$ is the closest vector in $C(A)$ to b

Hence $Ax = \hat{b}$ has a solution.

Let $\hat{x}$ be a solution to $Ax = \hat{b}$.

$\hat{x}$ is called a least-squares solution to $Ax = b$.

We already know how to find $\hat{x}$:

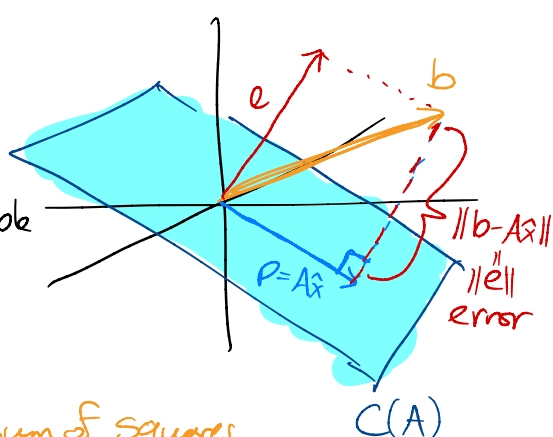
$$\text{Solve} \quad A^T A \hat{x} = A^T b$$

Warning: A^TA may not be invertible since A may not have linearly independent columns. So $\hat{x}$ need not be unique.

Geometrically

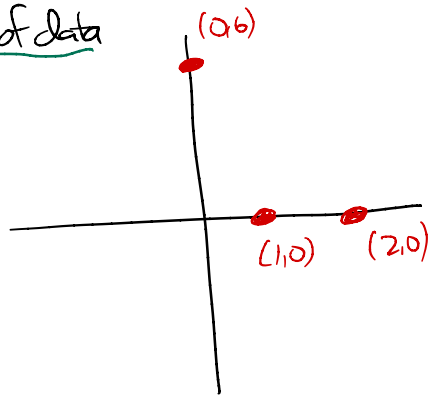
The solution $\hat{x} \in \mathbb{R}^n$ makes the error vector $e = b - Ax$ as small as possible

That is, $\hat{x}$ minimizes the value of $\|b - Ax\|$



this is a sum of squares
⇝ "least squares"

Special Case: least squares fitting of data



Find the line closest to the points (0,6), (1,0), and (2,0).

If a straight line $y = mx + b$ went through those points, then:

At (0,6), $m(0) + b = 6$
At (1,0), $m(1) + b = 0$
At (2,0), $m(2) + b = 0$

OR $\begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$.

Obviously, there is no solution since there is no such line. However, we can find a least-squares soln

$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}$, $A^T = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}$, $b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$

$A^TA = \begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix}$, $A^Tb = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$.

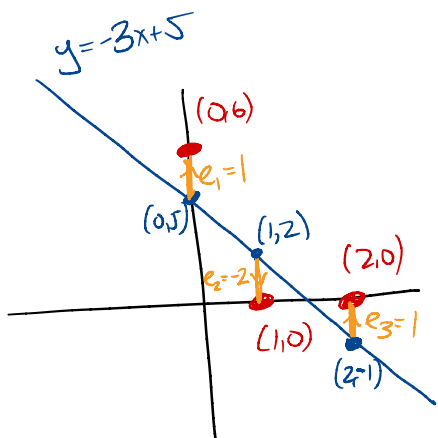
Solve $\begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix} \hat{x} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

$\left[\begin{array}{cc|c} 5 & 3 & 0 \\ 3 & 3 & 6 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{cc|c} 1 & 0 & -3 \\ 0 & 1 & 5 \end{array}\right]$

$\Rightarrow \begin{matrix} m = -3 \\ b = 5 \end{matrix}$

So the line $y = -3x + 5$ is the line of best fit.

(Can also use other functions)



Geometrically, $\hat{x}$ minimizes the length of the error vector

$e = b - A\hat{x} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$

the entries of e are the vertical distances from the line to the data points

So $\hat{x}$ minimizes the sum of squares of the vertical distances btwn data & line.

Orthonormal Bases and Gram-Schmidt

Def: The vectors $v_1, \dots, v_n \in \mathbb{R}^m$ are <u>orthogonal</u> if $v_i \cdot v_j = 0 \ \forall\, i \neq j$. If, in addition, $\|v_i\| = 1$ for all i, then they are called <u>orthonormal</u>.

Ex: $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ are orthonormal.

Key Calculation: If a matrix Q (not necessarily square) has orthonormal columns, then $Q^TQ = I$

(Q^T is the "left inverse" of Q)

proof: Let $Q = [q_1 \cdots q_p]$.

By definition, $q_i \cdot q_j = 0 \ \forall\, i \neq j$ and $q_i \cdot q_i = \|q_i\|^2 = 1 \ \forall\, i$.

Hence $Q^TQ = \begin{bmatrix} q_1^T \\ \vdots \\ q_p^T \end{bmatrix} [q_1 \cdots q_p]$

$$= \begin{bmatrix} q_1 \cdot q_1 & q_1 \cdot q_2 & \cdots & q_1 \cdot q_p \\ \vdots & & & \\ q_p \cdot q_1 & \cdots & & q_p \cdot q_p \end{bmatrix} = I$$

■

Def: A <u>square</u> matrix Q is called <u>orthogonal</u> if its columns are orthonormal.

(In this case $Q^TQ = I \implies Q^T = Q^{-1}$)

Ex: $Q = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$ is orthogonal.

Property: If Q is orthogonal, then $(Qx)\cdot(Qy) = x\cdot y$ for all $x, y \in \mathbb{R}^n$

Proof: $(Qx)\cdot(Qy) = (Qx)^T(Qy) = x^T Q^T Q y = x^T I y = x^T y = x\cdot y$ ▨

Two consequences:

- $\|Qx\|^2 = (Qx)\cdot(Qx) = x\cdot x = \|x\|^2$. So $\|Qx\| = \|x\|$.
- Let θ be the angle between x and y

  $(Qx)\cdot(Qy) = x\cdot y = \|x\|\|y\|\cos\theta = \|Qx\|\|Qy\|\cos\theta$

  $\Rightarrow$ θ is also the angle between Qx and Qy

Connection to linear transformations

Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation whose associated matrix Q is orthogonal.

Then by the above properties, T preserves length and angles:

i.e. $\|T(v)\| = \|v\|$ $\forall\, v \in \mathbb{R}^n$ and $\measuredangle(v,w) = \measuredangle(T(v), T(w))$

Ex: Let $Q = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$